

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Neural network target identification system for false alarm reduction

David Ye, Weston Edens, Thomas T. Lu, Tien-Hsin Chao

David Ye, Weston Edens, Thomas T. Lu, Tien-Hsin Chao, "Neural network target identification system for false alarm reduction," Proc. SPIE 7340, Optical Pattern Recognition XX, 73400K (13 April 2009); doi: 10.1117/12.820949

SPIE.

Event: SPIE Defense, Security, and Sensing, 2009, Orlando, Florida, United States

Neural network target identification system for false alarm reduction

David Ye^b, Weston Edens^c, Thomas T. Lu^{1a}, Tien-Hsin Chao^a

^aJet Propulsion Laboratory/CalTech, Pasadena CA, USA, ^bCalifornia Institute of Technology, Pasadena, CA, USA, ^cButler University/Purdue University, Indianapolis, IN, USA,

ABSTRACT

A multi-stage automated target recognition (ATR) system has been designed to perform computer vision tasks with adequate proficiency in mimicking human vision. The system is able to detect, identify, and track targets of interest. Potential regions of interest (ROIs) are first identified by the detection stage using an Optimum Trade-off Maximum Average Correlation Height (OT-MACH) filter combined with a wavelet transform. False positives are then eliminated by the verification stage using feature extraction methods in conjunction with neural networks. Feature extraction transforms the ROIs using filtering and binning algorithms to create feature vectors. A feed forward back propagation neural network (NN) is then trained to classify each feature vector and remove false positives. This paper discusses the test of the system performance and parameter optimizations process which adapts the system to various targets and datasets. The test results show that the system was successful in substantially reducing the false positive rate when tested on a sonar image dataset.

Key Words: automated target recognition, neural network, multi-stage ATR, false alarm reduction.

1. INTRODUCTION

The objective was to design an efficient computer vision system which can learn to detect multiple targets in large images with unknown backgrounds. Since the target size is small relative to the image size in this problem, there are many regions of the image which could potentially contain the target. A cursory analysis of every region can be computationally efficient, but may yield too many false positives. On the other hand, a detailed analysis of every region can yield better results, but may be computationally inefficient. JPL's multi-stage ATR system was designed to achieve an optimal balance between accuracy and computational efficiency by incorporating both models [1, 2]. As illustrated in Figure 1, the detection stage first identifies potential ROIs where the target may be present by performing a fast Fourier domain OT-MACH filter based correlation. Since threshold for this stage is chosen with the goal of detecting all true positives, a number of false positives are also detected as ROIs. The verification stage then transforms the regions of interest into feature space and eliminates false positives using an artificial neural network classifiers [3 - 5]. This paper focuses on the design test of the verification stage and its performance in eliminating false positives.

¹ e-mail: Thomas.T.Lu@jpl.nasa.gov , Tel: (818) 354-9513, Fax: (818) 393-4272

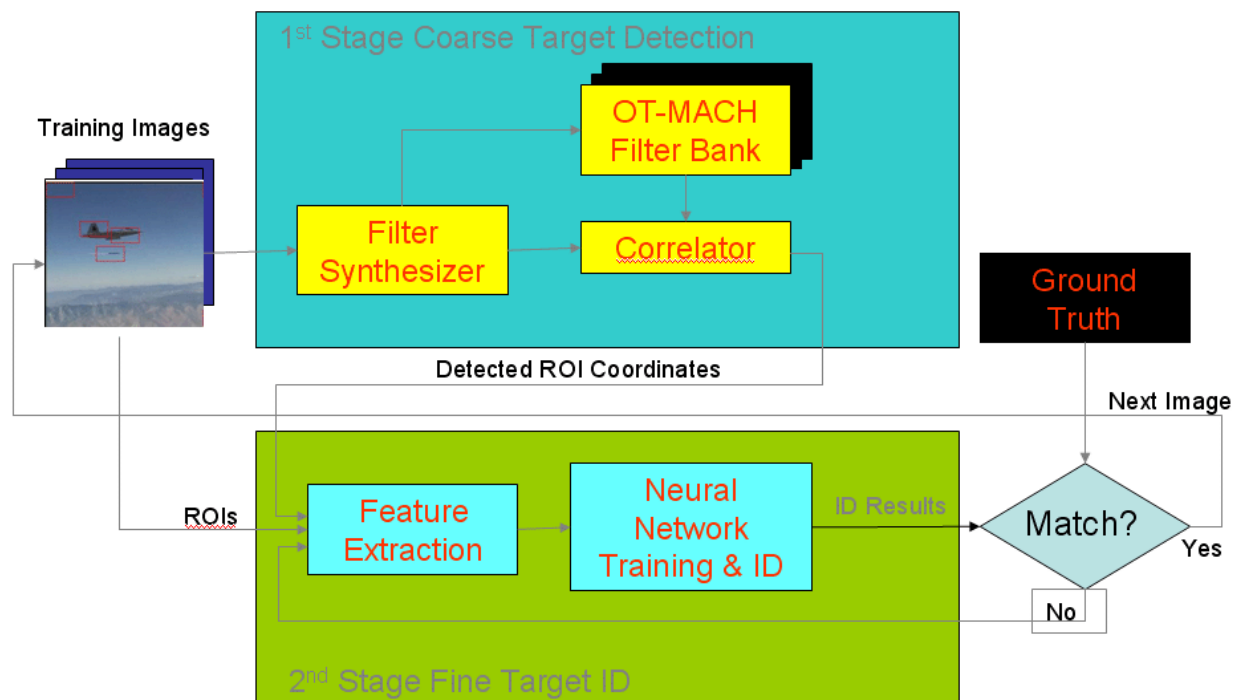


Figure 1: A multi-stage ATR system architecture.

A set of sonar image data containing underwater mines was used to evaluate the performance of the multi-stage ATR system. Four scrubbed sonar image quadrants are shown in Figure 2. These images contain high levels of background noise and the only distinguishing characteristic of the mines is a long narrow shadow. The images also contain debris and ocean crevices which cast shadows very similar to the mine shadows. The detection stage performed poorly on this dataset and a large number of false positives were detected. The verification stage must be able to eliminate most false positives without sacrificing detection rate in this dataset.

An artificial neural network was used to classify feature vectors as true or false positives. The standard feed-forward back-propagation variant was chosen for this task. Such neural networks have been successfully employed in image retrieval and object tracking systems [3, 4]. This computational model operates under the supervised learning paradigm. A set of input-output pairs were classified by hand to serve as the training set. Two separate neural networks were trained to classify feature vectors of the datasets as true or false positives. Cross validation was used to improve generalization.

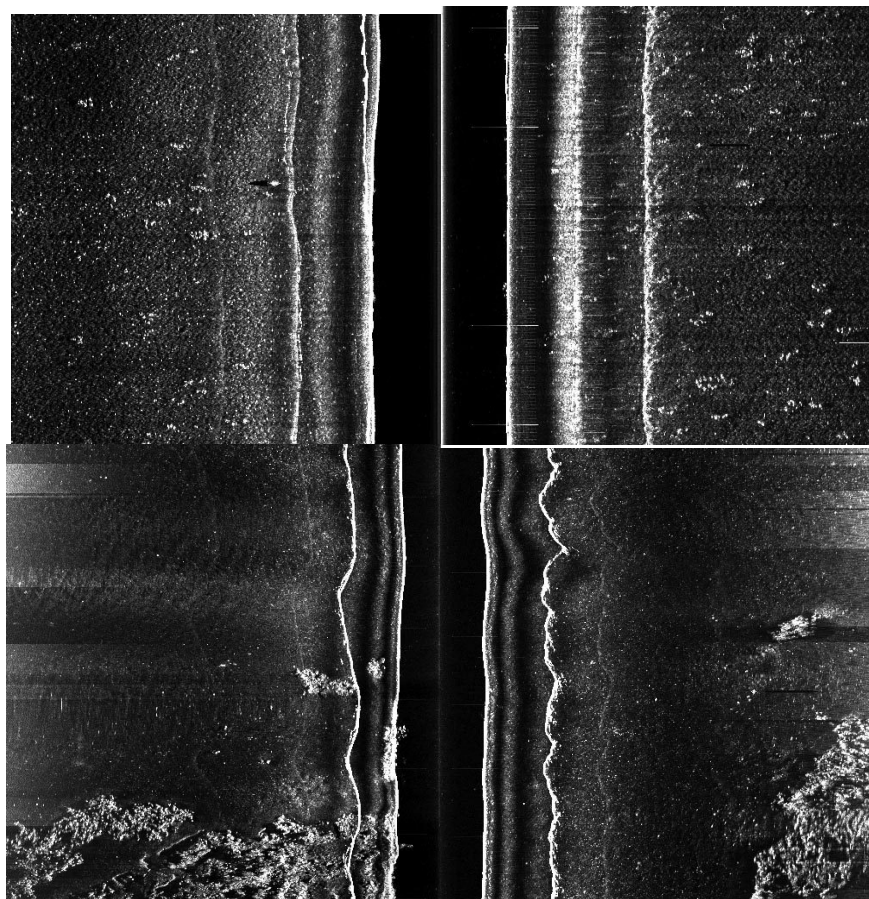


Figure 2: Sample Sonar Images (Scrubbed).

2. FIRST STAGE: OT-MACH CORRELATION FOR TARGET DETECTION

JPL has developed the detection stage using Grayscale Optical Correlation (GOC) and Neural Networks (NN) [1, 2]. Before the system can be implemented in hardware an algorithm must be created and optimized on a computer simulation. The GOC process uses an Optimum Trade-off Maximum Average Correlation Height (OT-MACH) filter, which performs in the Fourier domain finding correlations in the image that contain likely positions of targets. The OT-MACH filter is made from images of the target and depends on various measures: Output Noise Variance (ONV), Average Correlation Energy (ACE), Average Similarity Measure (ASM), and Average Correlation Height (ACH). The filter is created from the function: $E(h) = \alpha(\text{ONV}) + \beta(\text{ACE}) + \gamma(\text{ACH})$, [6]. The ACH term is insignificant so the parameters that govern the filter are α , β , and γ . These terms control properties of the filter such as noise and distortion tolerance, and peak shape; thus they must be optimized to create the best filter for a set of images. The OT-MACH filter looks for potential ROIs that pass to the next stage for identification and validation.

Since the targets could come with a variety of conditions, the first step was to classify the set of images into three categories: easy, medium, and hard, for further examination. The shape and contrast of the targets from the easy category provided a template for a set of training images from which the OT-MACH filter is created. Artificially

generated mine models were used to allow for better generalization. Each of the training images was 512x512 pixels with the target object centered in the image. Also, images were used to test the correlation of the filter to the target. For this process, the real image was cropped to fit the 512x512 window and the filter was applied to the test image to find the Peak height and PSR.

3. SECOND STAGE: NEURAL NETWORK FOR TARGET VERIFICATION

Once the OT-MACH filter has detected the likely ROIs that contain the target, the ROIs can go through more image processing and feature extraction to become inputs to a NN for verification. NN have various advantages that are ideal for pattern recognition; some advantages of biological NN that artificial ones mimic are parallelism, learning ability, generalization ability, adaptability, fault tolerance and low energy consumption². The NN must be trained on a specific training target set in order to have high accuracy. The combination of GOC and NNs provide speed and accuracy improvements to current ATR systems, allowing benefits for various applications [1].

3.1 Wavelet Transform

The second phase of filtering was added for the sonar images to further detect the mines that were missed by the OT-MACH filter. The wavelet transform breaks the image into different frequency components much like the Fourier transform. The 2-D wavelet transform for images splits the image according to different details in the image, which produces an image $\frac{1}{4}$ the size of the original. The image splits into Horizontal, Vertical, and Diagonal details. The 3rd level wavelet was used since it brought out the mine characteristics the best. The horizontal detail image is the most helpful since the mines have mainly horizontal characteristics. The wavelet method is very good at identifying the targets that stand out, but needs some improvement for hard to see targets or ones with low contrast. By combining the methods, a better detection rate is achieved but the false positive rate is increased [7]. Figure 3 shows the horizontal characteristic portion of the wavelet transform that is resized. The target has a good correlation map and can easily be detected.

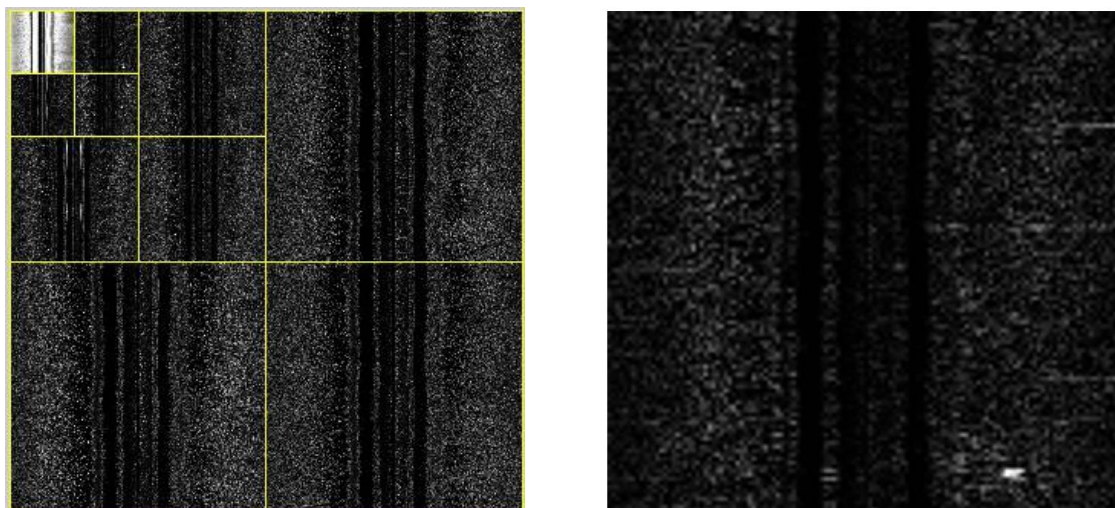


Figure 3: Image of mine after blur filter (left), Image of mine after blur filter and wavelet transform (right).

The next stage in the ATR process is a feature extraction method. The feature extraction method centers on horizontal and vertical binning of the ROI. A binning window size was chosen to be big enough to capture the entire target, but small enough to give a unique feature vector when compared to noise or false positives. Also, the number of pixels to bin is minimized so the input matrix to the NN is small. To help emphasize the characteristics of the target while reducing the background and noise in the image, the ROI was run through various filters. The final feature vector was formed by appending various filtered images of the ROI.

3.2 Feature Extraction

The ROIs identified by the detection stage represent small windows of the image which might contain a target. The dimensionality of these regions is the area of the window in pixel space. Even with small windows, this dimensionality is too large for input into a neural network. Much of the pixel data from each window also contains redundancies which lack substantial information about the target itself. Thus feature extraction techniques must be applied to transform the pixel data of each region into a smaller set of points which retains information about the target. Various feature extraction techniques have been described in literature and the choice of which to use is an important aspect to achieving optimal performance. Horizontal and vertical binning was used to provide features for the mine dataset. This relatively simple feature vector worked relatively well in reflecting the shadows of the mines.

Horizontal and vertical binning was chosen as the feature extraction method for the mine data [1]. This choice was motivated by the noisy, grayscale composition of the sonar images and hope that a simple method would be able to identify the distinguishing shadow of the target mines. As shown in Figure 4, the image windows representing the ROIs were binned into horizontal and vertical columns. The intensities along these bins were summed to obtain points of the feature vector. Feature vectors of true positives were characterized by two well defined dips representing areas of low intensity at the center of the horizontal and vertical bin profiles corresponding to the mine shadow, as illustrated in Figures 4(b) and (c). The edge feature is also binned, as shown in Figure 4(d). All 3 features are combined into a single vector for the neural network identification, as shown in Figure 4(e). Table 1 summarizes the feature extraction process.

The first step was to test various window sizes for the best fit for the targets. After trying various window sizes, the chosen window size was 56x24 pixels. This window size was big enough to capture most of the mines while minimizing the background size. When binned, this gives a vector of length 80. To reduce the size of the vector, the image is resized to a smaller window before binning so it contains all the features of the original image without as many pixels. This size of window was chosen to fit the entire mine into the window; the size is large compared to the mine, but some mines have different shapes. Since the mines had various shapes, sizes and contrasts, more than one image was binned for the feature vector to try to acquire as many features of the mine as possible. Many different filters were then tested to find the one that created the most unique feature vector for the targets. Two different edge filters were tested: the Prewitt and Sobel filters. These edge filters generated too much noise in the image when applied to the ROI so a blur filter was applied before the edge filter, but the resulting image was still too noisy to give a unique feature vector of the target. The next filter tested was a custom match filter. The general features of the mine were taken to produce the match filter with a light head and a long dark tail that contrasted with the background. A custom edge filter was also tested at this time. Of the custom filters the edge filter had the best correlation with the targets. Various blur filters and unsharp filters were also tested and the final feature vector is explained below.

The original ROI was first filtered with a blur filter and binned to make the first section of the feature vector. A custom edge filter was created and applied to the blurred image to append as the second part of the feature vector. The last section of the feature vector was the wavelet transform of the window. A sample feature vector is shown below. Statistical measures were also taken for the sonar feature windows. The feature vector of each window was processed, and ranges outside the acceptable level were set to zero.

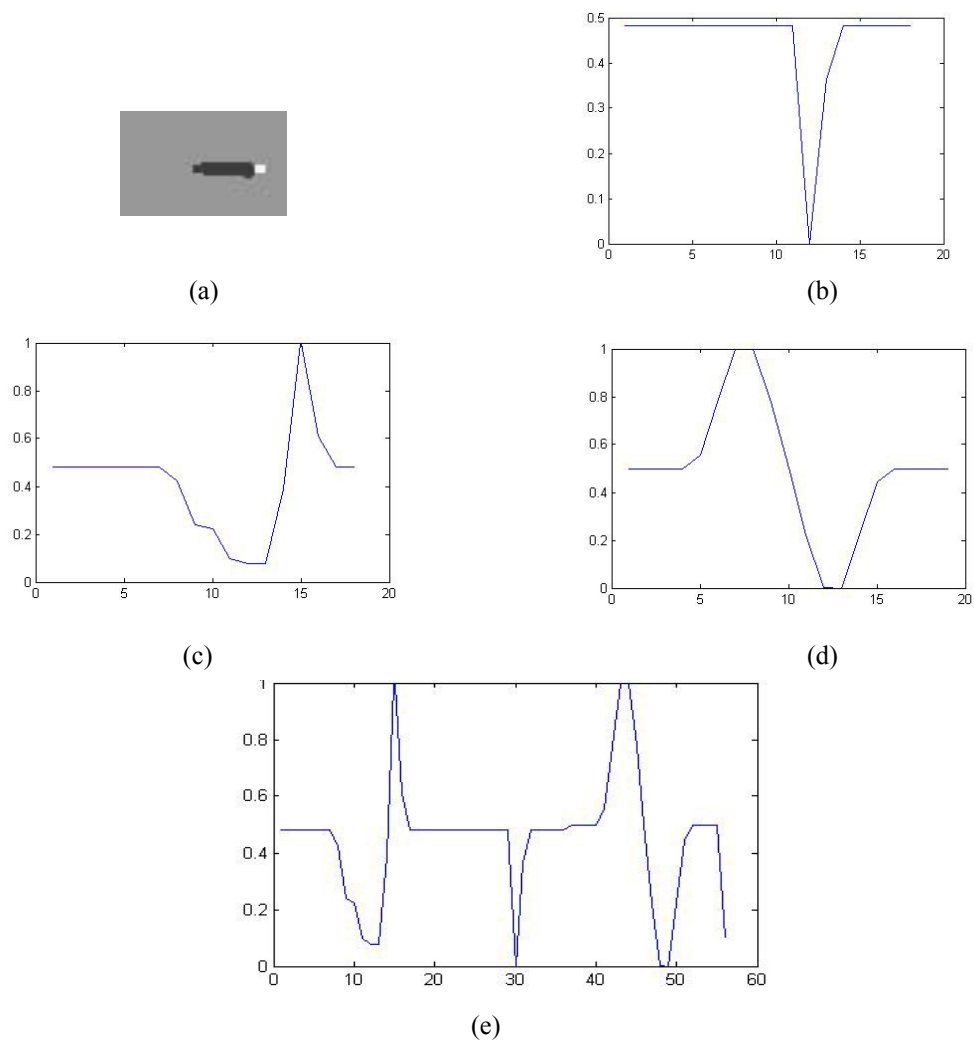


Figure 4: (a) A typical target model and its features: (a) vertical binning (b) horizontal binning; (c) vertical binning of edge enhanced image, and (d) combined feature for neural network training.

Table 1. Summary of feature extraction methods

Dataset	Mines
Window Size	48 x 24
Method	Intensity & Edge, Binning
Feature Size	18 Horizon, 18 Vertical, 20 edge, Total 56

3.3 Neural Network

An artificial neural network was used to classify feature vectors as true or false positives [1]. The standard feed-forward back-propagation variant was chosen for this task. Such neural networks have been successfully employed in image retrieval and object tracking systems. This computational model operates under the supervised learning paradigm. A set of input-output pairs were classified by hand to serve as the training set. Two separate neural networks were created and trained to classify feature vectors of the datasets as true or false positives. Cross validation was used to improve generalization. The trained neural networks substantially reduced the number of false positive per image from the detection stage results in both datasets. The false positive rate per image was reduced by 89% with a 91% detection rate in the boat dataset and reduced 97% with a 78% detection rate in the mine dataset.

An artificial neural network is an adaptive computational model inspired by the study of biological neural networks. The variant used in this system is the classic feed-forward back-propagation model. It is composed of separate layers of connected units called neurons. Every neuron of one layer is connected to every neuron of the next layer and each connection has an assigned weight value w . The output of a neuron y is calculated by a weighted sum of the inputs, x , into that neuron [3]:

$$y_j = f(X) = K(\text{Sum}(w_i x_i)) \quad (1)$$

where K is a transfer function which maps range between (0, 1).

The feature vector serves as the initial input into the neurons of the first hidden layer. The output of neurons for one layer then feeds into the neurons of the next layer until the output layer returns a value between [0, 1]. This architecture is known as feed-forward.

Since the connection weights are initially randomized, an untrained neural network models a completely random function. In order to classify data, it must learn from a training set which contains pairs of input, a feature vector, and output, either 0 or 1. We arbitrarily choose to denote false positives by 0 and true positives by 1. The back propagation algorithm performs gradient descent on the neural network to shift the weights in a direction which minimizes the error between the network and training set outputs using a performance metric such as mean square error.

The sonar training and test dataset consisted of approximately 100 images each containing one to two target mines. Mines were characterized primarily by the presence of their shadows. Due to the lack of many distinguishing features as well as the varying shapes and sizes of the shadows, the detection stage correlation filter detects many “mine-like” objects on this dataset and threshold was lowered to insure the initial detection of the mines. This resulted in 200 regions of interest per image of which the vast majority were false positives. A training set was designated by randomly selecting 30 images from the initial 60. This provided an initial set of 41 true positive and thousands of false positive regions of interest. The false positive regions were culled by only taking false positives from the first few images to prevent redundant vectors from being over expressed. The true positive regions were offset by a number of pixels in all four directions to improve generalization in the cases where regions were off center and repeated until a roughly equal number of true and false positive vectors were obtained. This resulted in a final training set of 1175 true positive and 1227 false positive regions.

The last part of the process is training the NN. A program was created to visualize the groups of positive and negative feature vectors to aid in training. After analysis of both groups, a set of positive and negative feature vectors was chosen for training. The training sets were split into three fractions: 70% was trained on using back-propagation, 15% was used for validation, and the remaining 15% was used for independent verification of the networks ability to generalize. The feature vector data was then input to a feed forward back propagation NN and simulations were run.

4. RESULTS AND DISCUSSION

The sonar image set was broken up into the four categories: easy, medium, hard and all images. For each set of images, we used a Frequency Relative Operating Characteristic (FROC) curve to test the performance of the system, plotting the average false positives per image against the True Positive Rate (TPR). The threshold of the neural network output was varied between 0 to 1, while measuring the TPR vs false alarms.

The set of easy images had the best performance, as expected. The points of interest on FROC curve are at 90% detected rate with 2.5 false positives per image and 96.5% detected with 10 false positives per image. The FROC curve in Figure 5 shows the performance of various NNs. The performance increases as training is carried out and better training sets are used.

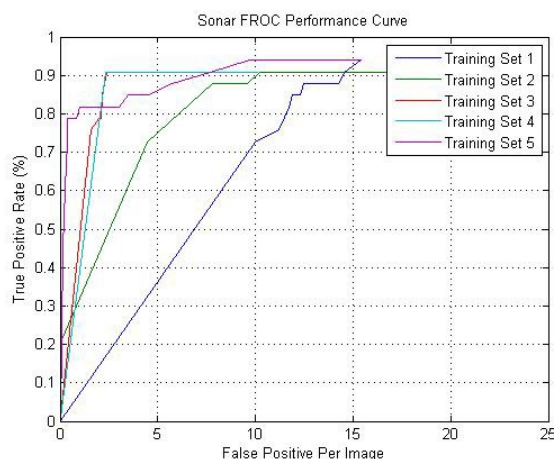


Figure 5: FROC curve of Sonar targets, Easy Set

The next set of images is the medium set of targets, which detect more false positives in order to have TPR above 90%. The targets in the medium set have less standard structure and more noise, which interferes with the NN training process. The FROC curve in Figure 6 also shows improvement in performance as training is carried out, but the overall performance is lower than the easy set. Some basic structures of the medium class of targets are in areas of background noise and low contrast which interfere with the characteristics of the targets. The resulting performance can be seen from the FROC curve and a point of interest on the curve is an 80% detection rate with 22 false positives per image, and 90% detection with 40 false positives per image.

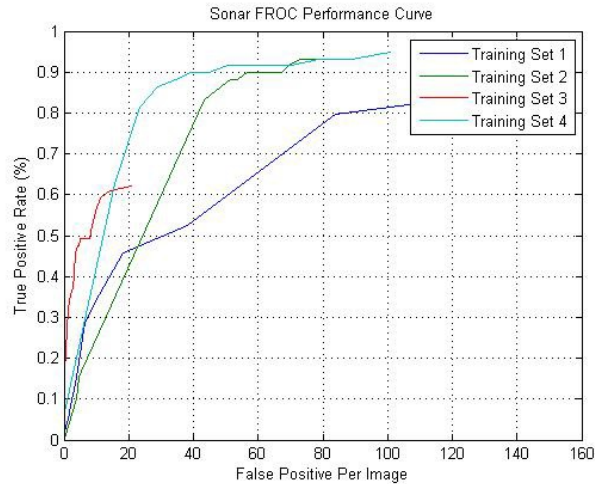


Figure 6: FROC curve of Sonar Mines, Medium Set

The set of images with the least commonality or structure is the hard set. The target feature structure was clouded by the background noise to a point that is almost unrecognizable by human eyes. Training the NN to search for a wide variety of features increases the false positive rate. As seen from the, the performance can be seen from the FROC curves in Figure 7. A point of interest on this FROC curve is 80% with 20 false positives per image.

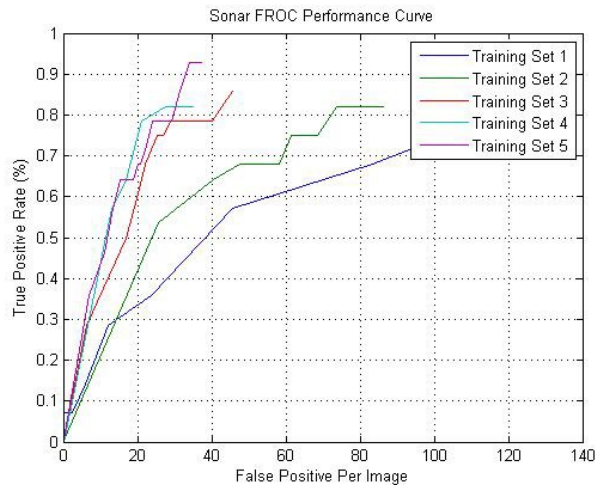


Figure 7: FROC curve of Sonar Mines, Hard Set

The set of all images had the most variation since it included all previous sets and needed to identify targets from all the sets while excluding many types of false positives. The performance is comparable to that of the medium and hard sets. Training the NN helped significantly as seen by the increase in performance from the FROC curve in Figure 8.

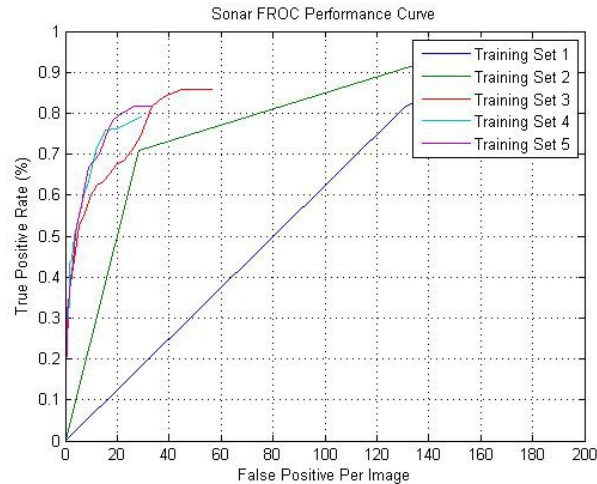


Figure 8: FROC curve of Sonar Mines, All Set

For each set, different NN training algorithms were tried to find the best performance. The FROC curve in Figure 9 shows the performance of the best four algorithms. The algorithms tested were cyclical order weight/bias, gradient decent, resilient back propagation, and the Levenberg-Marquardt training algorithm [3]. As seen from the plot below, the Levenberg-Marquardt algorithm has the best average performance.

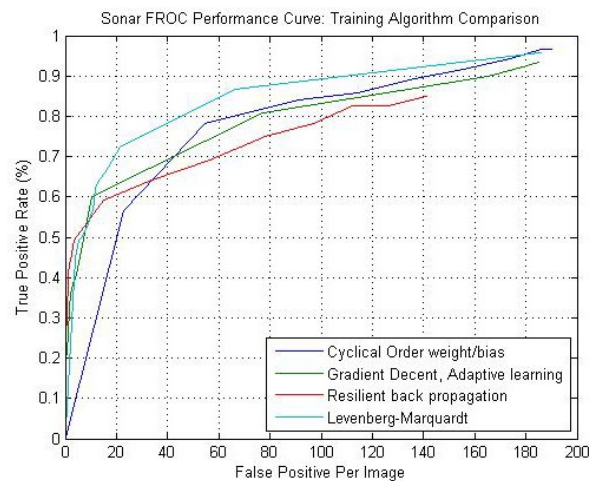


Figure 9: Comparing the performance of different NN learning algorithms.

CONCLUSIONS

The combination of the OT-MACH filter and the Neural Network is effective in ATR systems with appropriate optimizations. The trained neural networks substantially reduced the number of false positive per image from the detection stage results in both datasets. The false positive rate per image was reduced by 97% with over 80% detection rate in the sonar image dataset. To improve accuracy, targets with a wide variety of shapes require a more sophisticated approach to feature extraction techniques. The binning of filtered images did not give a unique enough feature vector for the Neural Network to completely discriminate targets from background features and the large

amounts of noise in the sonar images. We also showed that the process is generalizable to work with targets of varied structure.

Future work that could be carried out is to improve both the accuracy of the detection stage and the verification stage. The reduction of false positives in the detection stage will lead to improved accuracy in both the verification stage and overall performance. Another area for improvement is the reduction of computational time in the algorithm. The work was carried out by a simulation that is much slower than the system implemented in hardware, but making improvements in the algorithm would benefit the overall timing of the system. The major areas for improvement are the wavelet transform and the feature extraction phase, which are computationally expensive. The algorithm is set up for a very specific type and size of image. This could be made more adaptable for all types of images. A process that could be generalized to multiple images and targets more easily would be very beneficial.

ACKNOWLEDGMENTS

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration (NASA), and was sponsored under the SURF and USRP programs through NASA. The authors would like to acknowledge Mr. Tristan Winterhalter, Mr. Stephen Williams, and Mr. Oliver Johnson for useful help and discussions.

REFERENCES

1. Lu, T., Hughlett, C., Zhou, H., Chao, T. H., Hanan, J. C., "Neural network postprocessing of grayscale optical correlator," *Proc. SPIE* **5908**, *Optical Information Processing III*, 2005.
2. Hanan, J. C., Chao, T. H., Assad, C., Lu, T., Zhou, H., Hughlett, C., "Closed-Loop Automatic Target Recognition and Monitoring System," SPIE Conference, 2005.
3. Jain, A. K., Mao, J., & Mohiuddin, K. M., "Artificial Neural Networks: A Tutorial," 0018-9162/96 IEEE, March 1996.
4. Lu, T., Mintzer, D., "Hybrid neural networks for nonlinear pattern recognition," *Optical Pattern Recognition*, ed. by F. T. S. Yu & S. Jutamulia, Cambridge University Press, 1998.
5. Lu, T., Lerner, J., "Spectroscopy and hybrid neural network analysis," *Proceedings of the IEEE*, Volume 84, Issue 6, pp895 – 905, 1996.
6. Zhou, H., Hughlett, C., Hanan, J. C., Chao, T. H., "On the Development of Filter Management Module for Grayscale Optical Correlator," SPIE Conference, 2004.
7. Tagliarini, G. A., Page, E. W., Kesden, G. M., Chiang, D., McPartland, P. J., "Application of wavelet and neural processing to automatic target recognition," *SPIE Vol. 3069* pp 154-160, 1997.